

# Complexity of the relaxed Peaceman-Rachford splitting method for the sum of two maximal strongly monotone operators

Renato D.C. Monteiro\*, Chee-Khian Sim†

November 3, 2016

## Abstract

This paper considers the relaxed Peaceman-Rachford (PR) splitting method for finding an approximate solution of a monotone inclusion whose underlying operator consists of the sum of two maximal strongly monotone operators. Using general results obtained in the setting of a non-Euclidean hybrid proximal extragradient framework, convergence of the iterates, as well as pointwise and ergodic convergence rate, are established for the above method whenever an associated relaxation parameter is within a certain interval. An example is also discussed to demonstrate that the iterates may not converge when the relaxation parameter is outside this interval.

## 1 Introduction

In this paper, we consider the relaxed Peaceman-Rachford (PR) splitting method for solving the monotone inclusion

$$0 \in (A + B)(u) \quad (1)$$

where, for some  $\beta \geq 0$ ,  $A$  and  $B$  are maximal  $\beta$ -strongly monotone operators (with the convention that 0-strongly monotone means simply monotone). Recall that the relaxed PR splitting method is given by

$$x_k = x_{k-1} + \theta(J_B(2J_A(x_{k-1}) - x_{k-1}) - J_A(x_{k-1})),$$

where  $\theta > 0$  is a fixed relaxation parameter and  $J_T := (I + T)^{-1}$ . The special case of the relaxed PR splitting method in which  $\theta = 2$  is known as the Peaceman-Rachford (PR) splitting method and the one with  $\theta = 1$  is the widely-studied Douglas-Rachford splitting method (see for example [1, 6, 7]).

The analysis of the relaxed PR splitting method for the case in which  $\beta = 0$  has been undertaken in a number of papers. More specifically, convergence of the sequence of iterates generated by the relaxed PR splitting method is well-known when  $\theta < 2$  (see for example [1, 7]) and, according to [8], its limiting behavior for the case in which  $\theta \geq 2$  is not known. Actually, as a special case of the class of examples discussed in Section 5, it is shown that such sequence does not necessarily

---

\*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grant CMMI-1300221.

†Department of Mathematics, University of Portsmouth, Lion Gate Building, Lion Terrace, Portsmouth PO1 3HF. (email: chee-khian.sim@port.ac.uk). This research was made possible through support by Centre of Operational Research and Logistics, University of Portsmouth.

converge. An  $\mathcal{O}(1/\sqrt{k})$  (strong) pointwise convergence rate result is established in [10] for the relaxed PR splitting method when  $\theta \in (0, 2)$ . Moreover, when  $A = \partial f$  and  $B = \partial g$  where  $f$  and  $g$  are proper lower semi-continuous convex functions, [4, 5] derive strong pointwise (resp., ergodic) convergence rate bounds for the relaxed PR method when  $\theta \in (0, 2)$  (resp.,  $\theta \in (0, 2]$ ) under different assumptions on the functions. Assuming only  $\beta$ -strong monotonicity of  $A = \partial f$  and some smoothness property on  $f$ , [8] shows that the relaxed PR splitting method has linear convergence rate for  $\theta \in (0, 2 + \tau)$  for some  $\tau > 0$ . Finally, convergence rate results for special classes of  $A$  and  $B$  are derived in [3] for  $\theta$  lying in an interval of the latter type.

This paper considers the case in which  $\beta > 0$ , and hence the case when both  $A$  and  $B$  are strongly monotone. To the best of our knowledge, the analysis of the relaxed PR splitting method for solving (1) when  $\beta > 0$  is new. Unlike the analysis of [4, 5, 8] where  $A = \partial f$  and  $B = \partial g$ , this paper considers maximal  $\beta$ -strongly monotone point-to-set operators  $A$  and  $B$  which are not necessarily subdifferentials and do not necessarily satisfy any regularity assumptions. Strong pointwise and ergodic convergence rate results (Theorems 4.7 and 4.9) when  $\theta \in (0, 2 + \beta)$  and  $\theta \in (0, 2 + \beta]$ , respectively, and convergence of the iterates when  $\theta \in (0, 2 + \beta)$ , are established for the relaxed PR splitting method for every  $\beta \geq 0$ . As a consequence, convergence of the iterates generated by the PR splitting method is shown under the condition that  $\theta = 2$  and  $\beta > 0$ , which clearly contrasts to case  $\theta = 2$  and  $\beta = 0$  where non-convergence might occur (see Section 5).

Our analysis of the relaxed PR splitting method for solving (1) is based on viewing it as an inexact proximal point method, more specifically, as an instance of a non-Euclidean hybrid proximal extragradient (HPE) framework for solving the monotone inclusion problem. The proximal point method, proposed by Rockafellar [18], is a classical iterative scheme for solving the latter problem. Paper [19] introduces an Euclidean version of the HPE framework which is an inexact version of the proximal point method based on a certain relative error criterion. Iteration-complexities of the latter framework are established in [15] (see also [14]). Generalizations of the HPE framework to the non-Euclidean setting are studied in [9, 13, 20]. Applications of the HPE framework can be found for example in [11, 12, 14, 15].

This paper is organized as follows. Section 2 describes basic concepts and notation used in the paper. Section 3 discusses the non-Euclidean HPE framework which is applied to the study of the relaxed PR splitting method in Section 4. Finally, Section 5 provides an example showing that the iterates generated by the relaxed PR splitting method may not converge when  $\theta \geq 2(1 + \beta)$ .

## 2 Basic concepts and notation

This section presents some definitions, notation and terminology which will be used in the paper.

Let  $f$  and  $g$  be functions with the same domain and whose values are in set of positive real numbers. We write that  $f(\cdot) = \Omega(g(\cdot))$  if there exists constant  $K > 0$  such that  $f(\cdot) \geq Kg(\cdot)$ . Also, we write  $f(\cdot) = \Theta(g(\cdot))$  if  $f(\cdot) = \Omega(g(\cdot))$  and  $g(\cdot) = \Omega(f(\cdot))$ .

Let  $\mathcal{Z}$  be a finite-dimensional real vector space with inner product denoted by  $\langle \cdot, \cdot \rangle$  (an example of  $\mathcal{Z}$  is  $\mathbb{R}^n$  endowed with the standard inner product) and let  $\|\cdot\|$  denote an arbitrary seminorm in  $\mathcal{Z}$ . Its dual (extended) seminorm, denoted by  $\|\cdot\|_*$ , is defined as  $\|\cdot\|_* := \sup\{\langle \cdot, z \rangle : \|z\| \leq 1\}$ . The following result whose simple proof is omitted states some basic properties of the dual seminorm.

**Proposition 2.1** *Let  $A : \mathcal{Z} \rightarrow \mathcal{Z}$  be a self-adjoint positive semidefinite linear operator and consider the seminorm  $\|\cdot\|$  in  $\mathcal{Z}$  given by  $\|z\| = \langle Az, z \rangle^{1/2}$  for every  $z \in \mathcal{Z}$ . Then,  $\text{dom } \|\cdot\|_* = \text{Im}(A)$  and  $\|Az\|_* = \|z\|$  for every  $z \in \mathcal{Z}$ .*

Given a set-valued operator  $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ , its domain is denoted by  $\text{Dom}(T) := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}$  and its inverse operator  $T^{-1} : \mathcal{Z} \rightrightarrows \mathcal{Z}$  is given by  $T^{-1}(v) := \{z : v \in T(z)\}$ . The graph of  $T$  is defined by  $\text{Gr}(T) := \{(z, t) : t \in T(z)\}$ . The operator  $T$  is said to be monotone if

$$\langle z - z', t - t' \rangle \geq 0 \quad \forall (z, t), (z', t') \in \text{Gr}(T).$$

Moreover,  $T$  is maximal monotone if it is monotone and, additionally, if  $T'$  is a monotone operator such that  $T(z) \subset T'(z)$  for every  $z \in \mathcal{Z}$ , then  $T = T'$ . The sum  $T + T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$  of two set-valued operators  $T, T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$  is defined by  $(T + T')(z) := \{t + t' \in \mathcal{Z} : t \in T(z), t' \in T'(z)\}$  for every  $z \in \mathcal{Z}$ . Given a scalar  $\varepsilon \geq 0$ , the  $\varepsilon$ -enlargement  $T^{[\varepsilon]} : \mathcal{Z} \rightrightarrows \mathcal{Z}$  of a monotone operator  $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$  is defined as

$$T^{[\varepsilon]}(z) := \{t \in \mathcal{Z} : \langle t - t', z - z' \rangle \geq -\varepsilon, \quad \forall z' \in \mathcal{Z}, \forall t' \in T(z')\} \quad \forall z \in \mathcal{Z}. \quad (2)$$

### 3 A non-Euclidean hybrid proximal extragradient framework

This section discusses the non-Euclidean hybrid proximal extragradient (HPE) framework and its associated convergence and complexity results.

We start by introducing the definition of a distance generating function and its corresponding Bregman distance.

**Definition 3.1** *A proper lower semi-continuous convex function  $w : \mathcal{Z} \rightarrow [-\infty, \infty]$  is called a distance generating function if  $\text{int}(\text{dom } w) = \text{Dom}(\partial w) \neq \emptyset$  and  $w$  is continuously differentiable on this interior. Moreover,  $w$  induces the Bregman distance  $dw : \mathcal{Z} \times \text{int}(\text{dom } w) \rightarrow \mathbb{R}$  defined as*

$$(dw)(z'; z) := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \quad \forall (z', z) \in \mathcal{Z} \times \text{int}(\text{dom } w). \quad (3)$$

For simplicity, for every  $z \in \text{int}(\text{dom } w)$ , the function  $(dw)(\cdot; z)$  will be denoted by  $(dw)_z$  so that

$$(dw)_z(z') = (dw)(z'; z) \quad \forall (z', z) \in \mathcal{Z} \times \text{int}(\text{dom } w).$$

The following useful identities follow straightforwardly from (3):

$$\nabla(dw)_z(z') = -\nabla(dw)_{z'}(z) = \nabla w(z') - \nabla w(z) \quad \forall z, z' \in \text{int}(\text{dom } w), \quad (4)$$

$$(dw)_v(z') - (dw)_v(z) = \langle \nabla(dw)_v(z), z' - z \rangle + (dw)_z(z') \quad \forall z' \in \mathcal{Z}, \forall v, z \in \text{int}(\text{dom } w). \quad (5)$$

Our analysis of the non-Euclidean HPE framework presented in this section requires an extra property (first introduced in [9]) of the distance generating function, namely, that of being regular with respect to a seminorm.

**Definition 3.2** *Let distance generating function  $w : \mathcal{Z} \rightarrow [-\infty, \infty]$ , seminorm  $\|\cdot\|$  in  $\mathcal{Z}$  and convex set  $Z \subset \text{int}(\text{dom } w)$  be given. For given positive constants  $m$  and  $M$ ,  $w$  is said to be  $(m, M)$ -regular with respect to  $(Z, \|\cdot\|)$  if*

$$(dw)_z(z') \geq \frac{m}{2} \|z - z'\|^2 \quad \forall z, z' \in Z, \quad (6)$$

$$\|\nabla w(z) - \nabla w(z')\|_* \leq M \|z - z'\| \quad \forall z, z' \in Z. \quad (7)$$

Note that if the seminorm in Definition 3.2 is a norm, then (6) implies that  $w$  is strongly convex, in which case the corresponding  $dw$  is said to be nondegenerate. However, since we are not necessarily assuming that  $\|\cdot\|$  is a norm, our approach includes the case in which  $w$  is not strongly convex, or equivalently,  $dw$  is degenerate.

We observe that iteration-complexities of non-Euclidean HPE frameworks have already been studied in [9] and [13]. However, our approach in this section is different from these two papers as follows. Paper [13] deals with distance generating functions  $w$  (see Definition 3.1) which are not necessarily regular in the sense of Definition 3.2. Since it considers a larger class of distance generating functions than the one considered in this work, the results obtained in [13] are more limited in scope, i.e., only an ergodic convergence rate result is obtained for operators with bounded feasible domains (or, more generally, for the case in which the sequence generated by the HPE framework is bounded). On the other hand, by introducing the class of distance generating functions  $w$  satisfying Definition 3.2, paper [9] analyzes the behavior of a HPE framework for solving inclusions whose operators are strongly monotone with respect to  $w$  and uses these results to solve (not necessarily strongly) monotone inclusion via regularization.

The following result gives some useful properties of regular distance generating functions.

**Lemma 3.3** *Let  $w : \mathcal{Z} \rightarrow [-\infty, \infty]$  be an  $(m, M)$ -regular distance generating function with respect to  $(Z, \|\cdot\|)$  as in Definition 3.2. Then, the following statements hold:*

(a) *for every  $z, z' \in Z$ , we have*

$$\|\nabla(dw)_{z'}(z)\|_*^2 \leq \frac{2M^2}{m} \min\{(dw)_z(z'), (dw)_{z'}(z)\}; \quad (8)$$

(b) *for every  $l \geq 1$  and  $z_0, z_1, \dots, z_l \in Z$ , we have*

$$(dw)_{z_0}(z_l) \leq \frac{lM}{m} \sum_{i=1}^l \min\{(dw)_{z_{i-1}}(z_i), (dw)_{z_i}(z_{i-1})\}. \quad (9)$$

**Proof:** (a) It is easy to see that (8) immediately follows from (4), (6) and (7).

(b) Using (3) and (7), it is easy to see that

$$(dw)_z(z') = w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \leq \frac{M}{2} \|z - z'\|^2 \quad \forall z, z' \in Z. \quad (10)$$

Then, we have

$$(dw)_{z_0}(z_l) \leq \frac{M}{2} \|z_l - z_0\|^2 \leq \frac{M}{2} \left( \sum_{i=1}^l \|z_i - z_{i-1}\| \right)^2 \leq \frac{lM}{2} \sum_{i=1}^l \|z_i - z_{i-1}\|^2$$

which, in view of (6), immediately implies (9). ■

Throughout this section, we assume that  $w : \mathcal{Z} \rightarrow [-\infty, \infty]$  is an  $(m, M)$ -regular distance generating function with respect to  $(Z, \|\cdot\|)$  where  $Z \subset \text{int}(\text{dom } w)$  is a closed convex set and  $\|\cdot\|$  is a seminorm in  $\mathcal{Z}$ . Our problem of interest in this section is the monotone inclusion problem (MIP) for a maximal monotone operator

$$0 \in T(z) \quad (11)$$

where  $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$  is a maximal monotone operator and the following conditions hold:

**A0)**  $\text{Dom}(T) \subset Z$ ;

**A1)** the solution set  $T^{-1}(0)$  of (11) is nonempty.

We now state a non-Euclidean HPE (NE-HPE) framework for solving the MIP (11) which generalizes the ones studied in [15, 16, 19].

**Framework 1** (An NE-HPE variant for solving (11)).

(0) Let  $z_0 \in Z$  and  $\sigma \in [0, 1]$  be given, and set  $k = 1$ ;

(1) choose  $\lambda_k > 0$  and find  $(\tilde{z}_k, z_k, \varepsilon_k) \in Z \times Z \times \mathbb{R}_+$  such that

$$r_k := \frac{1}{\lambda_k} \nabla(dw)_{z_k}(z_{k-1}) \in T^{[\varepsilon_k]}(\tilde{z}_k), \quad (12)$$

$$(dw)_{z_k}(\tilde{z}_k) + \lambda_k \varepsilon_k \leq \sigma(dw)_{z_{k-1}}(\tilde{z}_k); \quad (13)$$

(2) set  $k \leftarrow k + 1$  and go to step 1.

**end**

We now make some remarks about Framework 1. First, it does not specify how to find  $\lambda_k$  and  $(\tilde{z}_k, z_k, \varepsilon_k)$  satisfying (12) and (13). The particular scheme for computing  $\lambda_k$  and  $(\tilde{z}_k, z_k, \varepsilon_k)$  will depend on the instance of the framework under consideration and the properties of the operator  $T$ . Second, if  $w$  is strongly convex on  $Z$  and  $\sigma = 0$ , then (13) implies that  $\varepsilon_k = 0$  and  $z_k = \tilde{z}_k$  for every  $k$ , and hence that  $r_k \in T(z_k)$  in view of (12). Therefore, the HPE error conditions (12)-(13) can be viewed as a relaxation of an iteration of the exact non-Euclidean proximal point method, namely,

$$0 \in \frac{1}{\lambda_k} \nabla(dw)_{z_{k-1}}(z_k) + T(z_k).$$

Third, if  $w$  is strongly convex on  $Z$ , then it can be shown that the above inclusion has a unique solution  $z_k$ , and hence that, for any given  $\lambda_k > 0$ , there exists a triple  $(\tilde{z}_k, z_k, \varepsilon_k)$  of the form  $(z_k, z_k, 0)$  satisfying (12)-(13) with  $\sigma = 0$ . Clearly, computing the triple in this (exact) manner is expensive, and hence computation of (inexact) quadruples satisfying the HPE (relative) error conditions with  $\sigma > 0$  is more computationally appealing. Fourth, even though the definition of a regular distance generating function does not exclude the case in which  $w$  is constant, such a case is not interesting from an algorithmic analysis point of view. In fact, if  $w$  is constant, then we have that  $\tilde{z}_1$  is already a solution of (11) since it follows from (13) with  $k = 1$  that  $\varepsilon_1 = 0$ , and hence that  $0 \in T(\tilde{z}_1)$  in view of (12) with  $k = 1$ .

**Lemma 3.4** *For every  $k \geq 1$ , the following statements hold:*

(a) *for every  $z \in \text{dom } w$ , we have*

$$(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) = (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle;$$

(b) *for every  $z \in \text{dom } w$ , we have*

$$(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) \geq (1 - \sigma)(dw)_{z_{k-1}}(\tilde{z}_k) + \lambda_k (\langle r_k, \tilde{z}_k - z \rangle + \varepsilon_k);$$

(c) for every  $z^* \in T^{-1}(0)$ , we have

$$(dw)_{z_{k-1}}(z^*) - (dw)_{z_k}(z^*) \geq (1 - \sigma)(dw)_{z_{k-1}}(\tilde{z}_k). \quad (14)$$

**Proof:** (a) Using (5) twice and using the definition of  $r_k$  given by (12), we obtain

$$\begin{aligned} (dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) &= (dw)_{z_{k-1}}(z_k) + \langle \nabla(dw)_{z_{k-1}}(z_k), z - z_k \rangle \\ &= (dw)_{z_{k-1}}(z_k) + \langle \nabla(dw)_{z_{k-1}}(z_k), \tilde{z}_k - z_k \rangle + \langle \nabla(dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle \\ &= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \langle \nabla(dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle \\ &= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle. \end{aligned}$$

(b) This statement follows as an immediate consequence of (a) and (13).

(c) This statement follows from (b), the fact that  $0 \in T(z^*)$  and  $r_k \in T^{[\varepsilon_k]}(\tilde{z}_k)$ , and the definition of  $T^{[\varepsilon]}(\cdot)$ .  $\blacksquare$

For the purpose of stating the results below, define

$$(dw)_0 = \inf\{(dw)_{z_0}(z^*) : z^* \in T^{-1}(0)\}. \quad (15)$$

It is not known whether the infimum in (15) can be attained.

**Lemma 3.5** For every  $i \geq 1$ , define

$$\theta_i = \max \left\{ \frac{\lambda_i^2 \|r_i\|_*^2}{\tau^2(1 + \sqrt{\sigma})^2}, \frac{\lambda_i \varepsilon_i}{\sigma} \right\},$$

where

$$\tau := \frac{\sqrt{2}M}{\sqrt{m}}.$$

Then,

$$(1 - \sigma) \sum_{i=1}^k \theta_i \leq (dw)_0.$$

**Proof:** For every  $i \geq 1$ , we have

$$\begin{aligned} \lambda_i \|r_i\|_* &= \|\nabla(dw)_{z_{i-1}}(\tilde{z}_i) - \nabla(dw)_{z_i}(\tilde{z}_i)\|_* \leq \|\nabla(dw)_{z_{i-1}}(\tilde{z}_i)\|_* + \|\nabla(dw)_{z_i}(\tilde{z}_i)\|_* \\ &\leq \tau \left[ (dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} + (dw)_{z_i}(\tilde{z}_i)^{1/2} \right] \leq \tau(1 + \sqrt{\sigma})(dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} \end{aligned}$$

The last inequality, (13) and the definition of  $\theta_i$  then imply that  $\theta_i \leq (dw)_{z_{i-1}}(\tilde{z}_i)$  for every  $i \geq 1$ . Hence, if  $z^* \in T^{-1}(0)$ , it follows from Lemma 3.4(c) that

$$(1 - \sigma) \sum_{i=1}^k \theta_i \leq (1 - \sigma) \sum_{i=1}^k (dw)_{z_{i-1}}(\tilde{z}_i) \leq (dw)_{z_0}(z^*) - (dw)_{z_k}(z^*) \leq (dw)_{z_0}(z^*).$$

The lemma now follows from the latter inequality and the definition of  $(dw)_0$ .  $\blacksquare$

**Lemma 3.6** Assume that  $\sigma < 1$ . Then, for every  $\alpha \in \mathbb{R}$  and every  $k \geq 1$ , there exists an  $i \leq k$  such that

$$\|r_i\|_* \leq \tau(1 + \sqrt{\sigma}) \sqrt{\frac{(dw)_0}{1 - \sigma} \left( \frac{\lambda_i^{\alpha-2}}{\sum_{j=1}^k \lambda_j^\alpha} \right)}, \quad \varepsilon_i \leq \frac{\sigma(dw)_0}{1 - \sigma} \left( \frac{\lambda_i^{\alpha-1}}{\sum_{j=1}^k \lambda_j^\alpha} \right). \quad (16)$$

**Proof:** It is easy to verify that the conclusion of the lemma is equivalent to the condition

$$\min_{i=1,\dots,k} \left\{ \frac{\theta_i}{\lambda_i^\alpha} \right\} \left( \sum_{i=1}^k \lambda_i^\alpha \right) \leq \frac{(dw)_0}{1-\sigma}$$

which in turn follows as an easy consequence of Lemma 3.5.  $\blacksquare$

**Theorem 3.7 (Pointwise convergence)** *If  $\sigma < 1$ , then the following statements hold:*

(a) *if  $\underline{\lambda} := \inf \lambda_k > 0$ , then for every  $k \in \mathbb{N}$  there exists  $i \leq k$  such that*

$$\|r_i\|_* \leq \tau(1 + \sqrt{\sigma}) \sqrt{\frac{(dw)_0}{1-\sigma} \left( \frac{\underline{\lambda}^{-1}}{\sum_{j=1}^k \lambda_j} \right)} \leq \frac{\tau(1 + \sqrt{\sigma})}{\underline{\lambda}\sqrt{k}} \sqrt{\frac{(dw)_0}{1-\sigma}}$$

$$\varepsilon_i \leq \frac{\sigma(dw)_0}{1-\sigma} \frac{1}{\sum_{i=1}^k \lambda_i} \leq \frac{\sigma(dw)_0}{(1-\sigma)\underline{\lambda}k},$$

(b) *for every  $k \in \mathbb{N}$ , there exists an index  $i \leq k$  such that*

$$\|r_i\|_* \leq \tau(1 + \sqrt{\sigma}) \sqrt{\frac{(dw)_0}{1-\sigma} \left( \frac{1}{\sum_{j=1}^k \lambda_j^2} \right)}, \quad \varepsilon_i \leq \frac{\sigma(dw)_0 \lambda_i}{(1-\sigma) \sum_{j=1}^k \lambda_j^2}. \quad (17)$$

**Proof:** Statements (a) and (b) follow from Lemma 3.6 with  $\alpha = 1$  and  $\alpha = 2$ , respectively.  $\blacksquare$

Due to the degeneracy of the Bregman distance  $dw$ , it is not possible to establish that the sequences  $\{\tilde{z}_k\}$  and  $\{z_k\}$  converge. However, under mild conditions on  $\{\lambda_k\}$ , Proposition 3.9 below shows that  $\{(dw)_{\tilde{z}_k}(z^*)\}$  and  $\{(dw)_{z_k}(z^*)\}$  both converge to 0 for some  $z^* \in T^{-1}(0)$ . Before stating this proposition, we first give the following technical lemma.

**Lemma 3.8** *Assume that  $\lim_{k \rightarrow \infty} (dw)_{z_{k-1}}(\tilde{z}_k) = 0$  and  $z^* \in T^{-1}(0)$  satisfies  $\liminf_{k \rightarrow \infty} (dw)_{\tilde{z}_k}(z^*) = 0$ . Then,*

$$\lim_{k \rightarrow \infty} (dw)_{\tilde{z}_k}(z^*) = 0, \quad \lim_{k \rightarrow \infty} (dw)_{z_k}(z^*) = 0. \quad (18)$$

**Proof:** Since  $\sigma \in [0, 1]$  and  $z^* \in T^{-1}(0)$ , it follows from Lemma 3.4(c) that  $\{(dw)_{z_k}(z^*)\}$  is non-increasing. Since by Lemma 3.3(b) with  $l = 2$ , we have

$$(dw)_{z_{k-1}}(z^*) \leq \frac{2M}{m} [(dw)_{z_{k-1}}(\tilde{z}_k) + (dw)_{\tilde{z}_k}(z^*)],$$

we conclude that from the assumptions of the lemma that  $\liminf_{k \rightarrow \infty} (dw)_{z_{k-1}}(z^*) = 0$ . The above two remarks then imply that  $\lim_{k \rightarrow \infty} (dw)_{z_k}(z^*) = 0$ . Using this conclusion together with the inequality

$$(dw)_{\tilde{z}_k}(z^*) \leq \frac{2M}{m} [(dw)_{z_{k-1}}(z^*) + (dw)_{z_{k-1}}(\tilde{z}_k)],$$

we conclude that  $\lim_{k \rightarrow \infty} (dw)_{\tilde{z}_k}(z^*) = 0$ .  $\blacksquare$

**Proposition 3.9** *Assume that  $\sigma < 1$ ,  $\sum_{i=1}^\infty \lambda_k^2 = \infty$  and  $\{\tilde{z}_k\}$  is bounded. Then, there exists  $z^* \in T^{-1}(0)$  such that (18) holds.*

**Proof:** The assumption that  $\sigma < 1$  and  $\sum_{i=1}^{\infty} \lambda_k^2 = \infty$  together with Theorem 3.8(b) imply that there exists subsequence  $\{(r_k, \varepsilon_k)\}_{k \in \mathcal{K}}$  converging to zero. Now, let  $z^*$  be an accumulation point of  $\{\tilde{z}_k\}_{k \in \mathcal{K}}$ . Then, by passing the inclusion  $r_k \in T^{\varepsilon_k}(\tilde{z}_k)$  to the limit, we conclude that  $0 \in T^0(z^*) = T(z^*)$ , and hence that  $z^* \in T^{-1}(0)$ . Hence, Lemma 3.4(c) and the assumption that  $\sigma < 1$  imply that  $\lim_{k \rightarrow \infty} (dw)_{z_{k-1}}(\tilde{z}_k) = 0$ . Clearly, (10) and the fact that  $z^*$  is an accumulation point of  $\{\tilde{z}_k\}_{k \in \mathcal{K}}$  imply that  $\liminf_{k \rightarrow \infty} (dw)_{\tilde{z}_k}(z^*) = 0$ . The conclusion of the proposition now follows directly from the previous lemma.  $\blacksquare$

From now on, we focus on the ergodic convergence rate of the NE-HPE framework. For  $k \geq 1$ , define  $\Lambda_k := \sum_{i=1}^k \lambda_i$  and the ergodic sequences

$$\tilde{z}_k^a = \frac{1}{\Lambda_k} \sum_{i=1}^k \lambda_i \tilde{z}_i, \quad r_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^k \lambda_i r_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^k \lambda_i (\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_k^a \rangle). \quad (19)$$

The following result derives the convergence rate of the above ergodic sequences.

**Theorem 3.10 (Ergodic convergence)** *For every  $k \geq 1$ , we have*

$$\varepsilon_k^a \geq 0, \quad r_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$$

and

$$\|r_k^a\|_* \leq \frac{2\tau\sqrt{(dw)_0}}{\Lambda_k}, \quad \varepsilon_k^a \leq \left(\frac{3M}{m}\right) \frac{2(dw)_0 + \rho_k}{\Lambda_k}$$

where

$$\rho_k := \max_{i=1, \dots, k} (dw)_{z_i}(\tilde{z}_i).$$

Moreover, the sequence  $\{\rho_k\}$  is bounded under either one of the following situations:

(a)  $\sigma < 1$ , in which case

$$\rho_k \leq \frac{\sigma(dw)_0}{1 - \sigma}; \quad (20)$$

(b)  $\text{Dom } T$  is bounded, in which case

$$\rho_k \leq \frac{2M}{m} [(dw)_0 + D] \quad (21)$$

where  $D := \sup\{\min\{(dw)_y(y'), (dw)_{y'}(y)\} : y, y' \in \text{Dom } T\}$  is the diameter of  $\text{Dom } T$  with respect to  $dw$ .

**Proof:** Let  $z^* \in T^{-1}(0)$  be given. Using (19), (12) and (4), we easily see that

$$\Lambda_k r_k^a = \nabla(dw)_{z_0}(z_k) = \nabla(dw)_{z_0}(z^*) - \nabla(dw)_{z_k}(z^*)$$

which, together with (8) and (14), imply that

$$\begin{aligned} \Lambda_k \|r_k^a\|_* &\leq \|\nabla(dw)_{z_0}(z^*)\|_* + \|\nabla(dw)_{z_k}(z^*)\|_* \\ &\leq \tau[(dw)_{z_0}(z^*)^{1/2} + (dw)_{z_k}(z^*)^{1/2}] \leq 2\tau(dw)_{z_0}(z^*)^{1/2}. \end{aligned}$$



This inequality together with definition of  $(dw)_0$  clearly imply the bound on  $\|r_k\|_*$ . To show the bound on  $\varepsilon_k^a$ , first note that Lemma 3.4(b) implies that for every  $z \in \text{dom } w$ ,

$$(dw)_{z_0}(z) - (dw)_{z_k}(z) \geq (1 - \sigma) \sum_{i=1}^k (dw)_{z_{i-1}}(\tilde{z}_i) + \sum_{i=1}^k \lambda_i (\langle r_i, \tilde{z}_i - z \rangle + \varepsilon_i).$$

Letting  $z = \tilde{z}_k^a$  in the last inequality and using the fact that  $(dw)_{z_0}(\cdot)$  is convex and  $\sigma \leq 1$ , we conclude that

$$\max_{i=1, \dots, k} (dw)_{z_0}(\tilde{z}_i) \geq (dw)_{z_0}(\tilde{z}_k^a) \geq \sum_{i=1}^k \lambda_i (\langle r_i, \tilde{z}_i - \tilde{z}_k^a \rangle + \varepsilon_i) = \Lambda_k \varepsilon_k^a$$

where the last equality is due to (19). On the other hand, (9) with  $l = 3$  implies that for every  $i \geq 1$  and  $z^* \in T^{-1}(0)$ ,

$$\begin{aligned} (dw)_{z_0}(\tilde{z}_i) &\leq \frac{3M}{m} [(dw)_{z_i}(\tilde{z}_i) + (dw)_{z_i}(z^*) + (dw)_{z_0}(z^*)] \\ &\leq \frac{3M}{m} [(dw)_{z_i}(\tilde{z}_i) + 2(dw)_{z_0}(z^*)] \end{aligned}$$

where the last inequality is due to Lemma 3.4(c). Combining the above two relations and using the definitions of  $\rho_k$  and  $(dw)_0$ , we then conclude that the bound on  $\varepsilon_k^a$  holds.

We now establish the bounds on  $\rho_k$  under either one of the conditions (a) or (b). First, if  $\sigma < 1$ , then it follows from (13) and Lemma 3.4(c) that  $(1 - \sigma)(dw)_{z_i}(\tilde{z}_i) \leq \sigma(1 - \sigma)(dw)_{z_{i-1}}(\tilde{z}_i) \leq \sigma(dw)_{z_{i-1}}(z^*) \leq \sigma(dw)_{z_0}(z^*)$  for every  $i \geq 1$  and  $z^* \in T^{-1}(0)$ , and hence that (20) holds. Assume now that  $\text{Dom } T$  is bounded. Then, for every  $i \geq 1$  and  $z^* \in T^{-1}(0)$ , we have

$$(dw)_{z_i}(\tilde{z}_i) \leq \frac{2M}{m} [(dw)_{z_i}(z^*) + \min\{(dw)_{\tilde{z}_i}(z^*), (dw)_{z^*}(\tilde{z}_i)\}] \leq \frac{2M}{m} [(dw)_{z_0}(z^*) + D]$$

which, in view of the definitions of  $\rho_k$  and  $(dw)_0$ , clearly implies that (21) holds. ■

## 4 The relaxed Peaceman-Rachford splitting method

This section studies the relaxed Peaceman-Rachford (PR) splitting method by applying results from the previous section.

For a given  $\beta > 0$ , an operator  $T : \mathcal{X} \rightrightarrows \mathcal{X}$  is said to be  $\beta$ -strongly monotone if

$$\langle w - w', x - x' \rangle \geq \beta \|x - x'\|_{\mathcal{X}}^2 \quad \forall (x, w), (x', w') \in \text{Gr}(T).$$

In what follows, we refer to monotone operators as 0-strongly monotone operators. This terminology has the benefit of allowing us to treat both the monotone and strongly monotone case simultaneously.

Throughout this section, we consider the monotone inclusion (1) where  $A, B : \mathcal{X} \rightrightarrows \mathcal{X}$  satisfy the following assumptions:

B0) for some  $\beta \geq 0$ ,  $A$  and  $B$  are maximal  $\beta$ -strongly monotone operators;

B1) the solution set  $(A + B)^{-1}(0)$  is non-empty.

We start by observing that (1) is equivalent to solving the augmented inclusion

$$\begin{aligned} 0 &\in \gamma A(u) + u - x \\ 0 &\in \gamma B(v) + x - v \\ 0 &= u - v \end{aligned}$$

where  $\gamma > 0$  is an arbitrary scalar.

Observe the skew-symmetry of the linear part of the above operator. Another way of writing the above operator is as

$$\begin{aligned} 0 &\in \gamma A(u) + u - x \\ 0 &\in \gamma B(v) + v + x - 2u \\ 0 &= u - v \end{aligned}$$

Note that the first and second equations are equivalent to

$$u = u(x) := J_{\gamma A}(x), \quad v = v(x) := J_{\gamma B}(2u - x) = J_{\gamma B}(2J_{\gamma A}(x) - x) \quad (22)$$

so that the third equation reduces to

$$0 = u(x) - v(x) = J_{\gamma A}(x) - J_{\gamma B}(2J_{\gamma A}(x) - x).$$

The Douglas-Rachford (DR) splitting method is the iterative procedure  $x_k = x_{k-1} + v(x_{k-1}) - u(x_{k-1})$ ,  $k \geq 1$ , started from some  $x_0 \in \mathcal{X}$ . It is known that the Douglas-Rachford splitting method is an exact proximal point method for some maximal monotone operator [6, 7]. Hence convergence of iterates of the method is guaranteed. More generally, the relaxed Peaceman-Rachford (PR) splitting method with relaxation parameter  $\theta > 0$  iterates as

$$(u_k, v_k) := (u(x_{k-1}), v(x_{k-1})) \quad x_k = x_k^\theta := x_{k-1} + \theta(v_k - u_k) \quad \forall k \geq 1. \quad (23)$$

When  $\theta = 1$ , we have the Douglas-Rachford splitting method, while when  $\theta = 2$ , we have the Peaceman-Rachford splitting method. The relaxed PR splitting method is an iterative process in the  $x$ -space but another way of looking at it is as an iterative process in the whole  $(u, v, x)$ -space as we will now discuss.

We now introduce a monotone inclusion which plays an important role in our analysis. For every  $\tilde{\theta} > 0$ , define the linear map  $\mathcal{L}_{\tilde{\theta}} : \mathcal{X} \times \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X} \times \mathcal{X} \times \mathcal{X}$  as

$$\mathcal{L}_{\tilde{\theta}}(z) = \mathcal{L}_{\tilde{\theta}}(u, v, x) := \begin{bmatrix} (1 - \tilde{\theta})I & \tilde{\theta}I & -I \\ (\tilde{\theta} - 2)I & (1 - \tilde{\theta})I & I \\ I & -I & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ x \end{pmatrix} \quad (24)$$

and the operator  $C : \mathcal{X} \times \mathcal{X} \times \mathcal{X} \rightrightarrows \mathcal{X} \times \mathcal{X} \times \mathcal{X}$  as

$$C(z) = C(u, v, x) := A(u) \times B(v) \times \{0\}. \quad (25)$$

The following result relates the relaxed PR splitting method with the inclusion  $0 \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$  and the subsequent one shows that this inclusion is monotone.

**Lemma 4.1** For a given  $x_{k-1} \in \mathcal{X}$  and  $\tilde{\theta} > 0$ , define

$$z_k^{\tilde{\theta}} = (u_k, v_k, x_k^{\tilde{\theta}}) \quad (26)$$

where  $u_k, v_k$  and  $x_k^{\tilde{\theta}}$  are as in (23), and set

$$a_k := \frac{1}{\gamma}(x_{k-1} - u_k), \quad b_k := \frac{1}{\gamma}(2u_k - v_k - x_{k-1}). \quad (27)$$

Then, we have:

$$-(1 - \tilde{\theta})u_k - \tilde{\theta}v_k + x_k^{\tilde{\theta}} = \gamma a_k \in \gamma A(u_k), \quad (28)$$

$$(2 - \tilde{\theta})u_k - (1 - \tilde{\theta})v_k - x_k^{\tilde{\theta}} = \gamma b_k \in \gamma B(v_k). \quad (29)$$

As a consequence, we have

$$(0, 0, u_k - v_k) = \mathcal{L}_{\tilde{\theta}}(z_k^{\tilde{\theta}}) + \gamma c_k \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z_k^{\tilde{\theta}}) \quad (30)$$

where

$$c_k := (a_k, b_k, 0).$$

**Proof:** Using the definition of  $(u(\cdot), v(\cdot))$ , the definition of  $(u_k, v_k, x_k^{\tilde{\theta}})$  in (23), and the definitions of  $a_k$  and  $b_k$ , we easily see that the conclusion of the lemma follows. Clearly, (30) follows as an immediate consequence of the conclusion of the lemma, definitions (24) and (25), and the definition of  $c_k$ . ■

The following result gives sufficient conditions for the maximal monotonicity of  $\mathcal{L}_{\tilde{\theta}} + \gamma C$ .

**Proposition 4.2** Assume that  $A, B : \mathcal{X} \rightrightarrows \mathcal{X}$  satisfy B0. Then,  $\mathcal{L}_{\tilde{\theta}} + \gamma C$  is a maximal monotone operator on  $\mathcal{X} \times \mathcal{X} \times \mathcal{X}$  for every  $\tilde{\theta} \in (0, 1 + (\gamma\beta)/2]$ .

**Proof:** Assumption B0 imply that  $A = \beta I + A_0$  and  $B = \beta I + B_0$  for some maximal monotone operators  $A_0, B_0 : \mathcal{X} \rightrightarrows \mathcal{X}$ . Therefore,  $\mathcal{L}_{\tilde{\theta}} + \gamma C$  can be written as

$$\mathcal{L}_{\tilde{\theta}} + \gamma C = \tilde{\mathcal{L}}_{\tilde{\theta}} + \gamma \tilde{C}$$

where

$$\tilde{\mathcal{L}}_{\tilde{\theta}}(z) = \tilde{\mathcal{L}}_{\tilde{\theta}}(u, v, x) = \begin{bmatrix} (1 - \tilde{\theta})I + \gamma\beta I & \tilde{\theta}I & -I \\ (\tilde{\theta} - 2)I & (1 - \tilde{\theta})I + \gamma\beta I & I \\ I & -I & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ x \end{pmatrix}$$

and

$$\tilde{C}(z) = \tilde{C}(u, v, x) = A_0(u) \times B_0(v) \times \{0\}.$$

It is easy to see that for every  $z = (u, v, x) \in \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ ,

$$\begin{aligned} \langle z, \tilde{\mathcal{L}}_{\tilde{\theta}}(z) \rangle &= [(1 - \tilde{\theta}) + \gamma\beta] (\|u\|_{\mathcal{X}}^2 + \|v\|_{\mathcal{X}}^2) + 2(\tilde{\theta} - 1)\langle u, v \rangle \\ &= (1 - \tilde{\theta})\|u - v\|_{\mathcal{X}}^2 + \gamma\beta (\|u\|_{\mathcal{X}}^2 + \|v\|_{\mathcal{X}}^2). \end{aligned}$$

which can be easily seen to be non-negative for every  $\tilde{\theta} \in (0, 1 + (\gamma\beta)/2]$ . Hence,  $\tilde{\mathcal{L}}_{\tilde{\theta}}$  is a monotone linear map for any  $\tilde{\theta} \in (0, 1 + (\gamma\beta)/2]$ . The conclusion of the proposition then follows by noting that the sum of a monotone linear map and a maximal monotone operator is a maximal monotone operator [1, 17]. ■

Let us now denote

$$\theta_0 := 1 + (\gamma\beta)/2. \quad (31)$$

Note that the value of  $\theta_0$  depends on  $\gamma$  and  $\beta$ . Clearly,  $\theta_0 = 1$  when  $\beta = 0$ .

The following result shows that the relaxed PR splitting method with  $\theta \in (0, 2\theta_0]$  can be viewed as an inexact instance of the NE-HPE method with respect to the monotone inclusion  $0 \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)(z)$  where

$$\tilde{\theta}_0 := \min\{\theta, \theta_0\}. \quad (32)$$

**Lemma 4.3** *For any  $\theta > 0$ ,  $\mathcal{L}_{\tilde{\theta}_0} + \gamma C$  is maximal monotone and the solution set  $(\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$  is given by*

$$\begin{aligned} (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0) &= \{(u^*, u^*, x^*) : \gamma^{-1}(x^* - u^*) \in A(u^*) \cap (-B(u^*))\} \\ &= \{(u^*, u^*, u^* + \gamma a^*) : a^* \in A(u^*), -a^* \in B(u^*)\}. \end{aligned}$$

As a consequence, if  $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$ , then  $u^* \in (A+B)^{-1}(0)$  and  $u^* = J_{\gamma A}(x^*)$ .

**Proof:** The conclusion of the lemma follows immediately from Proposition 4.2, the definitions of  $\mathcal{L}_{\tilde{\theta}_0}$  and  $\tilde{\theta}_0$  in (24) and (32), respectively, and some simple algebraic manipulation. ■

**Proposition 4.4** *Consider the sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated according to the relaxed PR splitting method (23) with any  $\theta > 0$  and define the sequences  $\{\tilde{z}_k = (\tilde{u}_k, \tilde{v}_k, \tilde{x}_k)\}$ ,  $\{\varepsilon_k\}$  and  $\{\lambda_k\}$  as*

$$\tilde{u}_k := u_k, \quad \tilde{v}_k := v_k, \quad \tilde{x}_k := x_{k-1} + \tilde{\theta}_0(v_k - u_k), \quad \varepsilon_k := 0, \quad \lambda_k := 1.$$

where  $\tilde{\theta}_0$  is defined above. Consider also the (degenerate) distance generating function given by

$$w(z) = w(u, v, x) = \frac{\|x\|_{\mathcal{X}}^2}{2\theta} \quad \forall z = (u, v, x) \in \mathcal{X} \times \mathcal{X} \times \mathcal{X}. \quad (33)$$

Then, the following statements hold:

(a) for every  $k \geq 1$ ,  $\{(z_k, \tilde{z}_k, \varepsilon_k)\}$  satisfies

$$r_k := \nabla(dw)_{z_k}(z_{k-1}) = (0, 0, u_k - v_k) = \mathcal{L}_{\tilde{\theta}_0}(\tilde{z}_k) + \gamma c_k \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)(\tilde{z}_k);$$

and hence  $\{(z_k, \tilde{z}_k, \varepsilon_k)\}$  satisfies (12) with  $T = \mathcal{L}_{\tilde{\theta}_0} + \gamma C$ ;

(b) the sequence  $\{(z_k, \tilde{z}_k, \varepsilon_k)\}$  satisfies (13) with  $\sigma = (\theta/\tilde{\theta}_0 - 1)^2$  and  $w$  as in (33).

As a consequence, the relaxed PR splitting method with  $\theta \in (0, 2\theta_0)$  (resp.,  $\theta = 2\theta_0$ ) is an NE-HPE instance with respect to the monotone inclusion  $0 \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)(z)$  in which  $\sigma < 1$  (resp.,  $\sigma = 1$ ).

**Proof:** First note that (26) and the definition of  $\tilde{z}_k$  imply that  $\tilde{z}_k = z_k^{\tilde{\theta}_0}$ . Hence, the second equality and the inclusion in (a) follow from (30) with  $\tilde{\theta} = \tilde{\theta}_0$ . The first equality in (a) follows immediately

from (23) and (33). Moreover, using (23), (33) and the definition of  $\tilde{x}_k$ , we conclude that for any  $\theta \in (0, 2\theta_0]$ ,

$$(dw)_{z_k}(\tilde{z}_k) = \frac{\|\tilde{x}_k - x_k\|_{\mathcal{X}}^2}{2\theta} = \left(\frac{\theta}{\tilde{\theta}_0} - 1\right)^2 \frac{\|\tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2}{2\theta} = \left(\frac{\theta}{\tilde{\theta}_0} - 1\right)^2 (dw)_{z_{k-1}}(\tilde{z}_k)$$

and hence that (13) is satisfied with  $\sigma = (1 - \theta/\tilde{\theta}_0)^2$ . Finally, the last conclusion follows from statements (a) and (b), and Lemma 4.3.  $\blacksquare$

We now make a remark about the special case of Proposition 4.4 in which  $\theta \in (0, \theta_0]$ . Indeed, in this case,  $\tilde{\theta}_0 = \theta$ , and hence  $\sigma = 0$  and  $\tilde{z}_k = z_k$  for every  $k \geq 1$ . Thus, the relaxed PR splitting method with  $\theta \in (0, \theta_0]$  can be viewed as an exact non-Euclidean proximal point method with distance generating function  $w$  as in (33) with respect to the monotone inclusion  $0 \in T(z) := (\mathcal{L}_\theta + \gamma C)(z)$ . Note also that the latter inclusion depends on  $\theta$ .

We will now show that sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated by the relaxed PR splitting method with  $\theta \in (0, 2\theta_0)$  converges, and also establish that this sequence and the sequence  $\{\tilde{z}_k = (u_k, v_k, \tilde{x}_k)\}$  (see Proposition 4.4) are bounded whenever  $\theta \in (0, 2\theta_0]$ .

**Theorem 4.5** *Consider the sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated by the relaxed PR splitting method with  $\theta \in (0, 2\theta_0]$  and the sequence  $\{\tilde{z}_k = (u_k, v_k, \tilde{x}_k)\}$  defined in Proposition 4.4. Then, the following statements hold:*

- (a)  $\{z_k\}$  and  $\{\tilde{z}_k\}$  are bounded;
- (b) if  $\theta < 2\theta_0$ , then  $\{z_k\}$  converges to some  $z^* \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$ , and hence both  $\{u_k\}$  and  $\{v_k\}$  converge to some  $u^* \in (A + B)^{-1}(0)$ .

**Proof:** (a) The assumption that  $\theta \in (0, 2\theta_0]$  together with the last conclusion of Proposition 4.4 imply that the relaxed PR splitting method is an NE-HPE instance. Hence, for any  $z^* \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$ , it follows from Lemma 3.4(c) that the sequence  $\{(dw)_{z_k}(z^*)\}$  is nonincreasing where  $w$  is the distance generating function given by (33). Clearly, this observation implies that  $\{x_k\}$  is bounded. This conclusion together with (22) and the nonexpansiveness of  $J_{\gamma A}, J_{\gamma B}$  imply that  $\{u_k\}$  and  $\{v_k\}$  are also bounded. Finally,  $\{\tilde{x}_k\}$  is bounded due to the definition of  $\tilde{x}_k$  in Proposition 4.4 and the boundedness of  $\{x_k\}$ ,  $\{u_k\}$  and  $\{v_k\}$ .

(b) When  $\theta \in (0, 2\theta_0)$ , it follows from Proposition 4.4 that the relaxed PR splitting method is an NE-HPE instance with  $\sigma < 1$  and  $\lambda_k = 1$  for all  $k$ . Also, (a) implies that  $\{\tilde{z}_k\}$  is bounded. Therefore, by Proposition 3.9, there exists  $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$  such that (18) holds where  $w$  is as in (33). Hence, we conclude that  $\{x_k\}$  converges to  $x^*$ . The second identity in (23) then imply that  $\{u_k - v_k\}$  converges to 0. Since  $J_{\gamma A}$  is continuous, it follows from the definition of  $u(x)$  in (22) and the first identity in (23) that  $\{u_k\}$  converges to  $J_{\gamma A}(x^*)$ , and hence to  $u^*$ , in view of Lemma 4.3. The latter two conclusions imply that  $\{v_k\}$  also converges to  $u^*$ . We have thus shown the first conclusion of (b). The second conclusion follows from the first one and Lemma 4.3.  $\blacksquare$

When  $\beta = 0$ , the above result shows that the sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated by the relaxed PR splitting method with  $\theta \in (0, 2)$  converges. The specific instance of Section 5 corresponding to  $\beta = 0$  shows that the sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated by the relaxed PR splitting method with  $\theta = 2$  does not necessarily converge. Note that when  $\beta > 0$  and  $\theta = 2$ , the latter sequence always converges in view of Theorem 4.5(b) and the fact that  $2\theta_0 = 2 + (\gamma\beta) > 2$ .

In the rest of this section, we consider the pointwise and ergodic convergence rate for the relaxed PR splitting method. We first endow the space  $\mathcal{Z} := \mathcal{X} \times \mathcal{X} \times \mathcal{X}$  with the semi-norm  $\|(u, v, x)\| := \|x\|_{\mathcal{X}}$  and hence Proposition 2.1 implies that

$$\|(0, 0, x)\|_* = \|x\|_{\mathcal{X}}. \quad (34)$$

It is also easy to see that the distance generating function  $w$  defined in (33) is  $(m, M)$ -regular with respect to  $(\mathcal{Z}, \|\cdot\|)$  where  $M = m = 1/\theta$  (see Definition 3.2).

Our next goal is to state a pointwise convergence rate bound for the relaxed PR splitting method. We start by stating a technical result which is well-known for the case where  $\beta = 0$  (see for example Lemma 2.4 of [10]). The proof for the general case, i.e.,  $\beta \geq 0$ , is similar and is given in the Appendix for the sake of completeness.

**Lemma 4.6** *Assume that  $\theta \in (0, 2\theta_0]$ . Then, for every  $k \geq 2$ , we have  $\|\Delta x_k\|_{\mathcal{X}} \leq \|\Delta x_{k-1}\|_{\mathcal{X}}$  where  $\Delta x_k := x_k - x_{k-1}$ .*

We now state the pointwise convergence rate result for the relaxed PR splitting method.

**Theorem 4.7** *Consider the sequence  $\{z_k = (u_k, v_k, x_k)\}$  generated by the relaxed PR splitting method with  $\theta \in (0, 2\theta_0)$ . Then, for every  $k \geq 1$  and  $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)^{-1}(0)$ ,*

$$a_k + b_k \in A(u_k) + B(v_k), \quad \gamma \|a_k + b_k\|_{\mathcal{X}} = \|u_k - v_k\|_{\mathcal{X}} \leq \frac{\sqrt{2}\|x_0 - x^*\|_{\mathcal{X}}}{\sqrt{k}\sqrt{2\tilde{\theta}_0 - \theta}}$$

where  $\|\cdot\|_{\mathcal{X}}$  denotes the inner product norm on  $\mathcal{X}$ .

**Proof:** The inclusion in the theorem follows from the two inclusions in (28) and (29). It follows from Proposition 4.4(a), and relations (27) and (33), that

$$r_k = \frac{1}{\theta}(0, 0, x_{k-1} - x_k) = (0, 0, u_k - v_k) = \gamma(0, 0, a_k + b_k) \quad (35)$$

and hence, in view of (34), that

$$\|r_k\|_* = \frac{1}{\theta}\|x_k - x_{k-1}\|_{\mathcal{X}} = \|u_k - v_k\|_{\mathcal{X}} = \gamma\|a_k + b_k\|_{\mathcal{X}}.$$

Since by Proposition 4.4, the relaxed PR splitting method with  $\theta \in (0, 2\theta_0)$  is an NE-HPE instance for solving the monotone inclusion  $0 \in (\mathcal{L}_{\tilde{\theta}_0} + \gamma C)(z)$  in which  $\sigma = (\theta/\tilde{\theta}_0 - 1)^2 < 1$  and  $\lambda_k = 1$  for all  $k \geq 1$ , it follows from the above relation, Lemma 4.6, Theorem 3.7, the fact that  $M = m = 1/\theta$ , and relations (15) and (33), that

$$\|r_k\|_* \leq \frac{\sqrt{2}M}{\sqrt{m}}(1 + \sqrt{\sigma})\sqrt{\frac{(dw)_0}{1 - \sigma} \left( \frac{1}{\sum_{j=1}^k \lambda_j^2} \right)} \leq \frac{\sqrt{2}\|x_0 - x^*\|_{\mathcal{X}}}{\sqrt{k}\sqrt{2\tilde{\theta}_0 - \theta}}.$$

The inequality of the theorem then follows from the above two observations.  $\blacksquare$

Our main goal in the remaining part of this section is to derive ergodic convergence rate bounds for the relaxed PR splitting method for any  $\theta \in (0, 2\theta_0]$ . We start by stating the following variation of the transportation lemma for maximal  $\beta$ -strongly monotone operators.

**Proposition 4.8** Assume that  $T$  is a maximal  $\beta$ -strongly monotone operator for some  $\beta \geq 0$ . Assume also that  $t_i \in T(u_i)$  for  $i = 1, \dots, k$ , and define

$$\bar{t}_k = \frac{1}{k} \sum_{i=1}^k t_i, \quad \bar{u}_k = \frac{1}{k} \sum_{i=1}^k u_i, \quad \varepsilon_k = \frac{1}{k} \sum_{i=1}^k \langle t_i - \beta u_i, u_i - \bar{u}_k \rangle \quad (36)$$

Then,  $\varepsilon_k \geq 0$  and  $\bar{t}_k \in T^{[\varepsilon_k]}(\bar{u}_k)$ .

**Proof:** The assumption that  $T$  is a maximal  $\beta$ -strongly monotone operator implies that  $T - \beta I$  is maximal monotone. Hence, it follows from the weak transportation formula (see Theorem 2.3 of [2]) applied to  $T - \beta I$  that  $\varepsilon_k \geq 0$  and  $\bar{t}_k - \beta \bar{u}_k \in (T - \beta I)^{[\varepsilon_k]}(\bar{u}_k)$ . The result then follows by observing that  $(T - \beta I)^{[\varepsilon_k]}(\bar{u}_k) + \beta \bar{u}_k \subseteq T^{[\varepsilon_k]}(\bar{u}_k)$ . ■

In order to state the ergodic iteration complexity bound for the relaxed PR splitting method, we introduce the ergodic sequences

$$\bar{u}_k = \frac{1}{k} \sum_{i=1}^k u_i, \quad \bar{v}_k = \frac{1}{k} \sum_{i=1}^k v_i, \quad \bar{a}_k = \frac{1}{k} \sum_{i=1}^k a_i, \quad \bar{b}_k = \frac{1}{k} \sum_{i=1}^k b_i \quad (37)$$

and the scalar sequences

$$\varepsilon'_k := \frac{1}{k} \sum_{i=1}^k \langle a_i - \beta u_i, u_i - \bar{u}_k \rangle, \quad \varepsilon''_k := \frac{1}{k} \sum_{i=1}^k \langle b_i - \beta v_i, v_i - \bar{v}_k \rangle.$$

**Theorem 4.9** Assume that  $\theta \in (0, 2\theta_0]$  and consider the ergodic sequences above. Then, for every  $k \geq 1$  and  $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\bar{\theta}_0} + \gamma C)^{-1}(0)$ ,

$$\bar{a}_k \in A^{[\varepsilon'_k]}(\bar{u}_k), \quad \bar{b}_k \in B^{[\varepsilon''_k]}(\bar{v}_k),$$

$$\gamma \|\bar{a}_k + \bar{b}_k\|_{\mathcal{X}} = \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} \leq \frac{2\|x_0 - x^*\|_{\mathcal{X}}}{k\theta}, \quad \varepsilon'_k + \varepsilon''_k \leq \frac{3(1 + 2(1 - \bar{\theta}_0/\theta)^2)\|x_0 - x^*\|_{\mathcal{X}}^2}{k\gamma\theta}$$

where  $\|\cdot\|_{\mathcal{X}}$  denotes the inner product norm on  $\mathcal{X}$ .

**Proof:** The first two inclusions follow from relations (28), (29) and (37), Assumption B0 and Proposition 4.8. We will now derive the two inequalities of the theorem using the fact that the PR splitting method with  $\theta \in (0, 2\theta_0]$  is an instance of the NE-HPE method. Letting  $r_k^a$  and  $\varepsilon_k^a$  be defined as in (19), we easily see from Proposition 4.4(a), and relations (35) and (37), that

$$r_k^a = (0, 0, \bar{u}_k - \bar{v}_k) = \gamma(0, 0, \bar{a}_k + \bar{b}_k)$$

and

$$\begin{aligned} \varepsilon_k^a &= \frac{1}{k} \sum_{i=1}^k \langle r_i, \tilde{z}_i - \tilde{z}_k^a \rangle = \frac{1}{k} \sum_{i=1}^k \langle \mathcal{L}_{\bar{\theta}_0}(\tilde{z}_i) + \gamma c_i, \tilde{z}_i - \tilde{z}_k^a \rangle \\ &\geq \frac{1}{k} \sum_{i=1}^k \gamma \langle c_i, \tilde{z}_i - \tilde{z}_k^a \rangle - \frac{\gamma\beta}{2k} \sum_{i=1}^k \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \\ &\geq \frac{1}{k} \sum_{i=1}^k \gamma (\langle a_i - \beta u_i, u_i - \bar{u}_k \rangle + \langle b_i - \beta v_i, v_i - \bar{v}_k \rangle) = \gamma(\varepsilon'_k + \varepsilon''_k), \end{aligned}$$

where in the first inequality we use the fact that

$$\begin{aligned} \sum_{i=1}^k \langle \mathcal{L}_{\tilde{\theta}_0}(\tilde{z}_i), \tilde{z}_i - \tilde{z}_k^a \rangle &= \sum_{i=1}^k \langle \mathcal{L}_{\tilde{\theta}_0}(\tilde{z}_i) - \mathcal{L}_{\tilde{\theta}_0}(\tilde{z}_k^a), \tilde{z}_i - \tilde{z}_k^a \rangle \geq (1 - \tilde{\theta}_0) \sum_{i=1}^k \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \\ &\geq (1 - \theta_0) \sum_{i=1}^k \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \end{aligned}$$

and the second inequality follows from

$$-\frac{1}{2} \sum_{i=1}^k \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \geq - \sum_{i=1}^k (\langle u_i, u_i - \bar{u}_k \rangle + \langle v_i, v_i - \bar{v}_k \rangle).$$

Hence, it follows from the definition of  $\|\cdot\|$ , the definition of  $w$  in (33), relations (15) and (34), and Theorem 3.10 with  $T = \mathcal{L}_{\tilde{\theta}_0} + \gamma C$ ,  $M = m = 1/\theta$  and  $\lambda_k = 1$  for all  $k$ , that

$$\gamma \|\bar{a}_k + \bar{b}_k\|_{\mathcal{X}} = \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} = \|r_k^a\|_* \leq \frac{2\sqrt{2}M(dw)_0^{1/2}}{\sqrt{m}\Lambda_k} \leq \frac{2\|x_0 - x^*\|_{\mathcal{X}}}{k\theta}$$

and

$$\gamma(\varepsilon'_k + \varepsilon''_k) \leq \varepsilon_k^a \leq \left(\frac{3M}{m}\right) \frac{2(dw)_0 + \rho_k}{\Lambda_k} \leq \frac{3(1 + 2(1 - \tilde{\theta}_0/\theta)^2)\|x_0 - x^*\|_{\mathcal{X}}^2}{k\theta}$$

since

$$\begin{aligned} \rho_k &:= \max_{i=1,\dots,k} (dw)_{z_i}(\tilde{z}_i) = \max_{i=1,\dots,k} \frac{\|x_i - \tilde{x}_i\|_{\mathcal{X}}^2}{2\theta} \\ &\leq (1 - \tilde{\theta}_0/\theta)^2 \max_{i=1,\dots,k} \frac{\|x_i - x_{i-1}\|_{\mathcal{X}}^2}{2\theta} \leq \frac{2(1 - \tilde{\theta}_0/\theta)^2\|x_0 - x^*\|_{\mathcal{X}}^2}{\theta}. \end{aligned}$$

■

We now make some remarks about the convergence rate bounds obtained in Theorem 4.9. In view of Lemma 4.3,  $x^*$  depends on  $\gamma$  according to

$$x^* = \gamma a^* + u^*, \quad a^* \in A(u^*) \cap -B(u^*).$$

Hence, letting

$$\begin{aligned} d_0 &:= \inf\{\|x_0 - u^*\|_{\mathcal{X}} : u^* \in (A + B)^{-1}(0)\}, \\ S &:= \sup\{\|a^*\| : a^* \in A(u^*) \cap -B(u^*), u^* \in (A + B)^{-1}(0)\}, \end{aligned}$$

and assuming that  $S < \infty$ , it is easy to see that Theorem 4.9 and (31) imply that the relaxed PR splitting method with  $\theta = 2\theta_0$  satisfies

$$\|\bar{a}_k + \bar{b}_k\|_{\mathcal{X}} \leq \frac{C_1(\gamma)}{\gamma k}, \quad \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} \leq \frac{C_1(\gamma)}{k}, \quad \varepsilon'_k + \varepsilon''_k \leq \frac{C_2(\gamma)}{k}$$

where

$$C_1(\gamma) = C_1(\gamma; \beta, d_0) = \Theta\left(\frac{d_0 + \gamma S}{1 + \beta\gamma}\right), \quad C_2(\gamma) = C_2(\gamma; \beta, d_0) = \Theta\left(\frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta\gamma)}\right).$$



When  $S/\beta \geq d_0$ , then  $\gamma = d_0/S$  minimizes both  $C_1(\cdot)$  and  $C_2(\cdot)$  up to a multiplicative constant, in which case  $C_1^* = \Theta(d_0)$ ,  $C_1^*/\gamma = \Theta(S)$  and  $C_2^* = \Theta(Sd_0)$  where

$$C_1^* = C_1^*(\beta, d_0) := \inf\{C_1(\gamma) : \gamma > 0\}, \quad C_2^* = C_2^*(\beta, d_0) := \inf\{C_2(\gamma) : \gamma > 0\}.$$

Note that this case includes the case in which  $\beta = 0$ . On the other hand, when  $S/\beta < d_0$ , then both  $C_1$  and  $C_2$  are minimized up to a multiplicative constant by any  $\gamma \geq d_0/S$ , in which case  $C_1^* = \Theta(S/\beta)$  and  $C_2^* = \Theta(S^2/\beta)$ . Clearly, in this case,  $C_1^*/\gamma$  converges to zero as  $\gamma$  tends to infinity.

Indeed, assume first that  $S/\beta \geq d_0$ . Then, up to some multiplicative constants, we have

$$C_1(\gamma) \geq \frac{d_0 + \gamma S}{1 + \beta\gamma} \geq \frac{d_0 + \gamma S}{1 + S\gamma/d_0} = d_0,$$

$$C_2(\gamma) \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta\gamma)} \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + S\gamma/d_0)} = \frac{d_0(d_0 + \gamma S)}{\gamma} = \frac{d_0^2}{\gamma} + Sd_0,$$

and hence that  $C_1^* = \Omega(d_0)$  and  $C_2^* = \Omega(Sd_0)$ . Moreover, if  $\gamma = d_0/S$ , then the assumption  $S/\beta \geq d_0$  implies that  $\beta\gamma \leq 1$ , and hence that  $C_1^* = \Theta(d_0)$  and  $C_2^* = \Theta(Sd_0)$ .

Assume now that  $S/\beta < d_0$ . Then, up to multiplicative constants, it is easy to see that

$$C_1(\gamma) \geq \frac{d_0 + \gamma S}{1 + \beta\gamma} \geq \frac{S}{\beta}$$

$$C_2(\gamma) \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta\gamma)} \geq \frac{(S/\beta + \gamma S)^2}{\gamma(1 + \beta\gamma)} = \frac{S^2}{\gamma\beta^2}(1 + \gamma\beta),$$

and hence that  $C_1^* = \Omega(S/\beta)$  and  $C_2^* = \Omega(S^2/\beta)$ . Moreover, if  $\gamma \geq d_0/S$ , then it is easy to see that  $C_1^* = \Theta(S/\beta)$  and  $C_2^* = \Theta(S^2/\beta)$ .

Based on the above discussion, the choice  $\gamma = d_0/S$  is optimal but has the disadvantage that  $d_0$  is generally difficult to compute. One possibility around this difficulty is to use  $\gamma = D_0/S$  where  $D_0$  is an upper bound on  $d_0$ .

## 5 An Example

By Theorem 4.5(b), the sequence  $\{x_k\}$  generated by the relaxed PR splitting method converges whenever  $\theta \in (0, 2 + \gamma\beta)$ . This section gives an instance of (1) for which the sequence  $\{x_k\}$  generated by the relaxed PR splitting method does not converge when  $\theta \geq 2(1 + \gamma\beta)$ .

Recall from (22) and (23) that the relaxed PR splitting method iterates as

$$x_{k+1} = x_k + \theta(J_{\gamma B}(2J_{\gamma A}(x_k) - x_k) - J_{\gamma A}(x_k)) \quad (38)$$

where  $\theta > 0$ . Without any loss of generality, we assume that  $\gamma = 1$  in (38).

We will now describe our instance. First, let  $A_0, B_0 : \mathbb{R} \rightrightarrows \mathbb{R}$  be defined as

$$\frac{1}{1+\beta}A_0(u) := \begin{cases} -\frac{\theta}{1+\beta} + \frac{1}{4}, & \text{if } u < -\frac{1}{4}, \\ \left[ -\frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k}\right) + \frac{1}{2k(k+1)}, -\frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k+1}\right) + \frac{1}{2(k+1)(k+2)} \right], & \text{if } u = -\frac{1}{2k(k+1)}, k = 1, \dots, \\ -\frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k+1}\right) + \frac{1}{2(k+1)(k+2)}, & \text{if } -\frac{1}{2k(k+1)} < u < -\frac{1}{2(k+1)(k+2)}, \\ & k = 1, \dots, \\ \left[ -\frac{\theta}{2(1+\beta)}, \frac{\theta}{2(1+\beta)} \right], & \text{if } u = 0, \\ \frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k+1}\right) - \frac{1}{2(k+1)(k+2)}, & \text{if } \frac{1}{2(k+1)(k+2)} < u < \frac{1}{2k(k+1)}, \\ & k = 1, \dots, \\ \left[ \frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k+1}\right) - \frac{1}{2(k+1)(k+2)}, \frac{\theta}{2(1+\beta)} \left(1 + \frac{1}{k}\right) - \frac{1}{2k(k+1)} \right], & \text{if } u = \frac{1}{2k(k+1)}, k = 1, \dots, \\ \frac{\theta}{1+\beta} - \frac{1}{4}, & \text{if } u > \frac{1}{4} \end{cases}$$

and

$$B_0(u) := \begin{cases} 2 - \theta + \frac{3\beta}{2}, & \text{if } u < -\frac{3}{2}, \\ \left[ \left(1 - \frac{\theta}{2}\right) \left(1 + \frac{1}{k}\right) + \beta \left(1 + \frac{1}{k+1}\right), \left(1 - \frac{\theta}{2}\right) \left(1 + \frac{1}{k+1}\right) + \beta \left(1 + \frac{1}{k+2}\right) \right], & \text{if } u = -1 - \frac{1}{k+1}, k = 1, \dots, \\ \left(1 - \frac{\theta}{2}\right) \left(1 + \frac{1}{k+1}\right) + \beta \left(1 + \frac{1}{k+2}\right), & \text{if } -1 - \frac{1}{k+1} < u < -1 - \frac{1}{k+2}, \\ & k = 1, \dots, \\ \left(\frac{\theta}{2} - 1 - \beta\right) u, & \text{if } -1 \leq u \leq 1, \\ \left(\frac{\theta}{2} - 1\right) \left(1 + \frac{1}{k+1}\right) - \beta \left(1 + \frac{1}{k+2}\right), & \text{if } 1 + \frac{1}{k+2} < u < 1 + \frac{1}{k+1}, \\ & k = 1, \dots, \\ \left[ \left(\frac{\theta}{2} - 1\right) \left(1 + \frac{1}{k+1}\right) - \beta \left(1 + \frac{1}{k+2}\right), \left(\frac{\theta}{2} - 1\right) \left(1 + \frac{1}{k}\right) - \beta \left(1 + \frac{1}{k+1}\right) \right], & \text{if } u = 1 + \frac{1}{k+1}, k = 1, \dots, \\ \theta - 2 - \frac{3\beta}{2}, & \text{if } u > \frac{3}{2} \end{cases}$$

Note that  $A_0$  and  $B_0$  are maximal monotone operators, where  $A_0$  and  $B_0$  are monotone since  $\theta \geq 2(1+\beta)$ . Letting  $A := A_0 + \beta I$  and  $B := B_0 + \beta I$  where  $\beta \geq 0$ , then it follows that  $A$  and  $B$  are maximal  $\beta$ -strongly monotone operators.

It is easy to see that

$$J_A(x) = J_{\frac{1}{1+\beta}A_0} \left( \frac{x}{1+\beta} \right), \quad J_B(x) = J_{\frac{1}{1+\beta}B_0} \left( \frac{x}{1+\beta} \right).$$

Hence, if  $x_k = \frac{\theta}{2} \left(1 + \frac{1}{k}\right)$ , it is easy to see that

$$J_A(x_k) = J_{\frac{1}{1+\beta}A_0} \left( \frac{x_k}{1+\beta} \right) = \frac{1}{2k(k+1)},$$

and hence that

$$J_B(2J_A(x_k) - x_k) = J_{\frac{1}{1+\beta}B_0} \left( \frac{2J_A(x_k) - x_k}{1+\beta} \right) = -1 - \frac{1}{k+1} = 2J_A(x_k) - \frac{2}{\theta}x_k.$$

Thus, (38) yields

$$\begin{aligned} x_{k+1} &= x_k + \theta[J_B(2J_A(x_k) - x_k) - J_A(x_k)] \\ &= x_k + \theta\left(2J_A(x_k) - \frac{2}{\theta}x_k - J_A(x_k)\right) \\ &= -x_k + \theta J_A(x_k) = -\frac{\theta}{2}\left(1 + \frac{1}{k+1}\right). \end{aligned}$$

Similarly, if  $x_k = -\frac{\theta}{2}\left(1 + \frac{1}{k}\right)$ , it is easy to check that

$$x_{k+1} = \frac{\theta}{2}\left(1 + \frac{1}{k+1}\right).$$

Hence, if  $x_0 = \theta$ , then the sequence  $\{x_k\}$  generated by the relaxed PR splitting method with  $\theta \geq 2(1 + \beta)$  for the above instance does not converge.

The above example shows that the sequence  $\{x_k\}$  does not converge whenever  $\theta \geq 2(1 + \gamma\beta)$ . On the other hand, Theorem 4.5(b) shows that  $\{x_k\}$  converges whenever  $0 < \theta < 2 + \gamma\beta$ . It is an open problem to investigate the behavior of  $\{x_k\}$  when  $\beta \in [2 + \gamma\beta, 2 + 2\gamma\beta)$  and  $\beta > 0$ . Finally, we mention without providing a proof that  $\{x_k\}$  converges when  $\theta = 2 + \gamma\beta$ ,  $\beta > 0$  and  $\mathcal{X} = \mathbb{R}$ .

## 6 Concluding remarks

This paper establishes convergence of the iterates and an  $\mathcal{O}(1/\sqrt{k})$  pointwise convergence rate bound for the relaxed PR splitting method for any  $\theta \in (0, 2 + \gamma\beta)$  by viewing it as an instance of a non-Euclidean HPE framework. It also establishes an ergodic  $\mathcal{O}(1/k)$  convergence rate bound for it for any  $\theta \in (0, 2 + \gamma\beta]$ . Furthermore, an example showing that its iterates do not necessarily converge for  $\theta \geq 2(1 + \gamma\beta)$  is given.

We observe that our analysis, in contrast to the ones in [5, 8], does not impose any regularity condition on  $A$  and  $B$  such as assuming one of them to be a Lipschitz, and hence point-to-point, operator. Also, if only one of the operators, say  $A$ , is assumed to be maximal  $\beta$ -strongly monotone, (1) is equivalent to  $0 \in (A' + B')(u)$  where  $A' := A - (\beta/2)I$  and  $B' := B + (\beta/2)I$  are now both  $(\beta/2)$ -strongly monotone. Thus, to solve (1), the relaxed PR method with  $(A, B)$  replaced by  $(A', B')$  can be applied, thereby ensuring convergence of the iterates, as well as pointwise and ergodic convergence rate bounds, for values of  $\theta > 2$ .

## References

- [1] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.
- [2] R.S. Burachik, C.A. Sagastizábal, and B.F. Svaiter.  $\varepsilon$ -enlargements of maximal monotone operators: Theory and applications. In *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods (Lausanne, 1997)*, volume 22 of *Applied Optimization*, pages 25–43. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
- [3] D. Davis. Convergence rate analysis of the forward-douglas-rachford splitting scheme. *SIAM Journal on Optimization*, 25:1760–1786, 2015.

- [4] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. *arXiv preprint arXiv:1406.4834v3*, 2015.
- [5] D. Davis and W. Yin. Faster convergence rates of relaxed peaceman-rachford and admm under regularity assumptions. *arXiv preprint arXiv:1407.5210v3*, 2015.
- [6] J. Eckstein and D.P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.
- [7] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problem, Volume II*. Springer-Verlag, New York, 2003.
- [8] P. Giselsson and S. Boyd. Linear convergence and metric selection for douglas-rachford splitting and admm. *arXiv preprint arXiv:1410.8479v5*, 2016.
- [9] M.L.N. Goncalves, J.G. Melo, and R.D.C. Monteiro. Improved pointwise iteration-complexity of a regularized admm and of a regularized non-euclidean hpe framework. *arXiv preprint arXiv:1601.01140v1*, 2016.
- [10] B. He and X. Yuan. On the convergence rate of douglas-rachford operator splitting method. *Mathematical Programming, Series A*, 153:715–722, 2015.
- [11] Y. He and R.D.C. Monteiro. Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player nash equilibrium problems. *SIAM Journal on Optimization*, 25:2182–2211, 2015.
- [12] Y. He and R.D.C. Monteiro. An accelerated hpe-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM Journal on Optimization*, 26:29–56, 2016.
- [13] O. Kolossoski and R.D.C. Monteiro. An accelerated non-euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. *Preprint*, 2015.
- [14] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of Tseng’s modified F-B splitting and korpelevich’s methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM J. Optim.*, 21(4):1688–1720, 2011.
- [15] R.D.C. Monteiro and B.F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20:2755–2787, 2010.
- [16] R.D.C. Monteiro and B.F. Svaiter. Complexity of variants of tseng’s modified f-b splitting and korpelevich’s methods of hemi-variational inequalities with applications to saddle point and convex optimization problems. *SIAM Journal on Optimization*, 21:1688–1720, 2011.
- [17] R.T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Transactions of the American Mathematical Society*, 149:75–88, 1970.
- [18] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

- [19] M.V. Solodov and B.F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued and Variational Analysis*, 7:323–345, 1999.
- [20] M.V. Solodov and B.F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of bregman functions. *Mathematics of Operations Research*, 25:214–230, 2000.

## Appendix

**Proof of Lemma 4.6:** To simplify notation, let

$$\Delta x = \Delta x_k, \quad \Delta x^- := \Delta x_{k-1}, \quad \Delta u = u_k - u_{k-1}, \quad \Delta v = v_k - v_{k-1}, \quad \Delta a = a_k - a_{k-1}, \quad \Delta b = b_k - b_{k-1}.$$

Then, it follows from the second identity in (23) and relation (27) that

$$\Delta x = \Delta x^- + \theta(\Delta v - \Delta u), \quad \gamma \Delta a = \Delta x^- - \Delta u, \quad \gamma \Delta b = 2\Delta u - \Delta v - \Delta x^-. \quad (39)$$

Also, the two inclusions in (28) and (29) together with the  $\beta$ -strong monotonicity of  $A$  and  $B$  imply that

$$\langle \Delta a, \Delta u \rangle \geq \beta \|\Delta u\|_{\mathcal{X}}^2, \quad \langle \Delta b, \Delta v \rangle \geq \beta \|\Delta v\|_{\mathcal{X}}^2.$$

Combining the last two identities in (39) with the above inequalities, we obtain

$$\langle \Delta x^- - \Delta u, \Delta u \rangle \geq \gamma \beta \|\Delta u\|_{\mathcal{X}}^2, \quad \langle 2\Delta u - \Delta v - \Delta x^-, \Delta v \rangle \geq \gamma \beta \|\Delta v\|_{\mathcal{X}}^2.$$

Adding these two last inequalities and simplifying the resulting expression, we obtain

$$\langle \Delta x^-, \Delta u - \Delta v \rangle + 2\langle \Delta u, \Delta v \rangle \geq (1 + \gamma \beta) [\|\Delta u\|_{\mathcal{X}}^2 + \|\Delta v\|_{\mathcal{X}}^2] \quad (40)$$

From the first equality in (39), we have

$$2\theta \langle \Delta x^-, \Delta u - \Delta v \rangle = \|\Delta x^-\|_{\mathcal{X}}^2 - \|\Delta x\|_{\mathcal{X}}^2 + \theta^2 \|\Delta v - \Delta u\|_{\mathcal{X}}^2,$$

which upon substituting into (40), the following is true:

$$\|\Delta x^-\|_{\mathcal{X}}^2 - \|\Delta x\|_{\mathcal{X}}^2 \geq 2\theta \left( \left[ 1 + \gamma \beta - \frac{\theta}{2} \right] (\|\Delta u\|_{\mathcal{X}}^2 + \|\Delta v\|_{\mathcal{X}}^2) + 2 \left[ \frac{\theta}{2} - 1 \right] \langle \Delta u, \Delta v \rangle \right).$$

Note that the right-hand side in the above inequality is greater than or equal to zero if  $\theta \in (0, 2\theta_0]$ . Hence, we have if  $\theta \in (0, 2\theta_0]$ ,

$$\|\Delta x\|_{\mathcal{X}} \leq \|\Delta x^-\|_{\mathcal{X}}.$$

■